DOCUMENT RESUME

ED 097 357 TM 003 986

AUTHOR	Pfaff, Judy K.; Schmidt, William H.
TITLE	Mastery Learning Strategies Applied to the Teaching of Statistics.
PUB DATE	[74]
NOTE	17p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th Chicago, Illinois, April 1974)
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Academic Achievement; Diagnostic Tests; *Graduate Students; *Learning Processes; Pretests; *Statistics; Teaching Methods
IDENTIFIERS	*Mastery Learning; Midterm Examinations

ABSTRACT

The purpose of this study is to describe and compare the performances of students in a beginning course in educational statistics. Modified mastery learning strategies such as the use of a pretest, diagnostic tests, parallel midterms, and an extended time allowance to master the content were employed during the course. Data was available on 359 graduate students over a 2-year period. The results of this nonexperimental study support the usefulness of the pretest; show that grades improve when more diagnostic tests are taken; scores almost always increase with the second parallel midterm; increased time is accompanied by decreased performance; masters students always do better than doctoral students; and the classes in which the mastery strategies were used were about the same overall as previous classes. (Author)

ED 097357

BEST COPY AVAILABLE

MASTERY LEARNING STRATEGIES APPLIED TO THE TEACHING OF STATISTICS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Judy K. Pfaff
and
William H. Schmidt
Michigan State Univ.

TM 003 986

BEST COPY AVAILABLE

In recent years there has been an ever increasing need for the educational researcher to develop basic statistical skills. An essential part of an educational program to meet these needs usually includes one or a sequence of courses in statistics. For many of the graduate students, this will be their first formal experience with a mathematical subject since entering college. Teaching practices in the initial encounters have to account for marked individual differences in background and interests, motivations, and mathematical competencies. For several years at Michigan State University a three sequence course in statistics has been given by the Educational Psychology Department to meet the needs of the various graduate programs in education and other social science fields. The, practices and empirical outcomes of the practices for the first course in this sequence, Education 869, will be described and compared here:

There are three general objectives for the course:

Students will learn:

1. About inferential statistics and the uses of these techniques in behavioral research contexts.
2. A series of statistical techniques and where they are applicable.
3. How to handle procedures (data and formulas) in a mechanical sense.

Only basic mathematical and algebraic skills are assumed at the beginning of the course. However, the content is sequential and cumulatively arranged such that the mastery of later topics depends upon the use or understanding of the earlier ones. The learning of statistics may be quite different in approach, ie, more mathematical and precise, than any recent course most students may have dealt with.

Thus the major problems encountered in teaching 869 are the diverse backgrounds and abilities of the students and the nature of the material to be taught. One set of instructional strategies for dealing with these

BEST COPY AVAILABLE

and similar problems has been developed, ie, the mastery learning model (Bloom, 1968). It assumes that all, or almost all students can learn given certain conditions and practices. The purpose of this paper, then, is to present certain modified aspects of the mastery learning model as applied to a beginning statistics course for graduate students and to show empirically the results of using these strategies over a two year time span.

Prior to registering for the course a pretest of algebraic skills must be taken. The results of this test are used to: 1) recommend a remedial or preparatory course to be taken prior to 869, 2) to indicate the extent to which skills must be relearned or reviewed and 3) to perhaps give a hint of the difficulty level or study time that might be needed to complete the course. Identification of the aptitudes for a given learning task associated with an individual's learning rate has been identified as one of the major research problems in the application of mastery learning strategies (Block, 1971).

The main instructive strategy consists of twenty two-hour lectures given over a ten week term. During the second year of this study, a packet of detailed lecture notes were available. These were designed to be used in class and are supposed to free the student from extensive note taking and allow the student to focus more attention on the immediate comprehension of the lecture material. All of the lectures are also available on cassette tapes which may be borrowed and used for additional study or to make up any missed lectures. They could be used to progress through the course at different rates, too. Three different texts are recommended for supplementary sources of explanation. Their appropriate sections are noted, and their timing is keyed to the lecture notes. A one hour small group problem session is held once a week. Students may ask to have any problems they

BEST COPY AVAILABLE

have had trouble with fully explained and solved in detail. These options can be considered components of varied instructional materials or alternate learning resources suggested by Bloom (Block, 1971).

Mastery conditions are fully defined prior to instruction in accordance with Bloom's preconditions for mastery learning (Block, 1971). Objectives, course content, and summative evaluation requirements, are explained during the first lecture and are available prior to registration upon request. A fixed grading standard was used for everyone. It was initially developed using the performance of three similar classes conducted in the year prior to this study. As the grade scale was adjusted slightly from year to year, the letter grade was felt unsuitable for use in comparisons across years. Percents served as a comparative variable instead.

Time allowed for learning has both instructional as well as university organizational problems (Block, 1971). The midterm, part of the summative evaluation, was offered upon arrangement. This resulted in half of the class opting to defer their grades and finish up the next term. In subsequent terms a deadline for taking the midterm, by the eighth week was imposed. A parallel midterm may be taken any time after the first midterm and before the final exam. The higher of the two midterms is used for calculating the grade. Tests taken by arrangement usually mean that the tests are taken alone or with a few others. There is a class time designated for the final exam. Those that feel that they are not yet ready, may choose to wait. They can revisit course lectures and attend problem sessions in subsequent terms and do what ever additional study they feel is necessary in preparation for their final exam. The students that eventually do take the final, but sometime after the term in which they were originally enrolled

BEST COPY AVAILABLE

will hereafter be called "deferred." Those students that have never finished or whose university imposed limit of 2 years was expended will be considered as "dropped." As with the final exam any notes, books and even calculators may be used as resource materials during the tests. Adequate time was allowed. Much effort was made to remove the usual tensions associated with testing in a mathematics subject area.

Formative nongraded evaluation procedures promoted by Bloom (Block, 1971) were used in 869. Twelve lesson units with separate answer sheets for self-study were given to all classes. Five diagnostic tests that cover all the material on the class tests are included in the packet for the second year of the study and the answers were available on request. In the first year, the same diagnostics were arranged and administered like the midterms. Students were urged to use the diagnostics as simulations of class tests. Students were not at any time, though, held accountable for the mastery or even the use of these lessons and diagnostics as would probably be the case in a stricter application of the mastery model.

A take home project of 4 to 6 problems provides an experience with more complete research problems. A special session on the use of desk calculators is held and students are encouraged to employ calculators in the completion of their projects. Several calculators are available for use during the course and in their later research experiences on campus. The take home project was not a requirement for two of the classes, summer sessions, included within this study.

Grades in 869 are assigned according to a predetermined scale. Percentages are calculated as a weighted combination of the highest midterm score, the project (terms that had one), and the final exam. A range of percentages on the predetermined scale is used to locate the number grade (0 to 4.0).

BEST COPY AVAILABLE

Two years of data covering 8 terms of ten week classes which were conducted using the above strategies has been collected. This was not an experimental study and cautions must be exercised in attempting to account for any differences found in analysis of these data. For the following standard conditions that were met for each of the classes, we do feel justified in comparing classes and students:

1. Same instructor.
2. Same instructional materials and similar lecture notes.
3. Same summative evaluation requirements.
4. Unreleased midterms and finals that were almost identical across terms. They might vary in removal of a bad item, change in numbers, or addition of a new item. (This study used percentages rather than raw scores for all of the analyses.)

Of 373 students who registered for the course in the two year period, 305 completed, 54 dropped, and 14 were removed from the analysis for various reasons. Those removed included undergraduates, visitors, and other atypical cases. The majority of the students are majoring in education and of those 45% are in a masters program while 51% are in a doctoral program. Entry characteristics including major, class (masters or doctoral), sex, residence (foreign versus native of U.S.), and last GPA (from bachelor's or master's work) were available on most but not all students. Frequencies for class, major, and sex are enumerated in tables 1-3 (appendix).

There are several issues within this program of instruction for beginning statistics which can be studied. Scores on two forms of the midterm, the project, the final exam, the final weighted percentage for assigning the grade, and the grades have been kept as they apply to all students. Diagnostic test scores are recorded for terms 1 through 3 and pretest scores for terms 5 through 8. The modified aspects of the mastery model that can be examined empirically within this nonexperimental study deal with the pretest, diagnostics, and a flexible time schedule. General comparisons can be made between those classes that had packets of notes versus own note taking and varied

BEST COPY AVAILABLE

learning resources versus the prior lecture and text classes. In addressing these issues regression and analysis of variance will be used. As is true of many nonexperimental studies, missing data has prevented examination of some interesting questions. Other aspects maybe speculated upon by comparing only frequencies and percentages. A description and discussion of the results and how they relate to those mastery model strategies as reported above will complete this paper.

The three main dependent variables for the study are the percentages on the midterm (highest of the two taken), a final exam, and the weighted percentage combining the midterm, final, and project. In further discussions these will be referred to as midterm, final, and percent, remember each is a percentage score. The lower of the two midterms will be labeled midterm I. Initial analysis with the project score showed that its variance was small, thus we decided not to use it as a separate dependent variable.

Though the pretest was in use for both years of the study only those pretests taken in the second year were recorded (terms 5-8, 123 subjects). The major question of interest to be answered concerning the pretest is: Was the pretest a useful tool for predicting final performance in the class? In other words, were the recommendations given to the students based on their pretest scores justified? A regression analysis (chi square test of association between dependent and independent variables with probability of significance less than .0001) shows that final performance can be predicted from the pretest score.

X^2	Independent Variable	Dependent Variable	R	R^2 (%)
$p < .0001$	Pretest	Midterm	.54	28.90
		Final	.58	33.26
		Percent	.56	31.65

BEST COPY AVAILABLE

An examination of the raw regression coefficients indicates that a change of 1 point on the pretest is accompanied by a change of approximately one percentage point on the midterm and the final percentage. Since many of the students score as low as 10 on the pretest an increase of 10 or 20 points on the pretest would predict a corresponding increase of 10% to 20% on the final.

Notice how the correlations increased between the pretest and midterm and the pretest and the final. Airasian correlated the ratings from a prior exposure inventory given students in a graduate class in an Introductory Testing Methods course using mastery learning strategies, with progress tests in the course (Psychology in the Schools, 1973). His correlations started at .60 and decreased to -.07 on the final test. He suggested that the correlations went down because appropriate feedback reduced the initial student differences. Lack of enough time during a ten week term to reduce the initial math background differences may be the problem here. Also because of the sequential nature of the material such differences if not eliminated will remain important throughout the course.

The diagnostics as described previously were administered as ungraded tests in terms 1 through 3. Many questions can be investigated about these conditions. First, do the number of diagnostics taken, Ndiags, anywhere from 0 to 5, predict final performance? A regression analysis using 115 subjects showed statistical significance ($p < .0443$).

χ^2	Independent Variable	Dependent Variable	R	R^2 (%)
$p < .0443$	Ndiags	Midterm	.20	3.84
		Final	.21	4.30
		Percent	.18	3.20

As the dependent variables were percentages, we tried several transformations, but the results were similar. The raw beta weights reveal an apparent change of 1 percentage point on the midterm and 2 percentage points on the final.

BEST COPY AVAILABLE

If the maximum number of diagnostics, 5 were taken instead of none of them, we could predict a corresponding ten percentage point increase on the final. This could result in a whole grade increase on the final summative evaluation. Causal interpretations have to be made cautiously, though, since students were under no obligation to take these diagnostics and, possibly, the brighter and better motivated students took more of the diagnostics in which case the Ndiags would be a surrogate measure for ability or motivation.

The first two diagnostics cover material that is tested on the midterm. A regression analysis to predict the combined scores on the first two diagnostics, DIAG 1 + 2, from the midterm, 95 subjects, gave the following results:

X^2	Independent Variable	Dependent Variable	R	R^2(%)
$p < .1117$	Midterm	Diag 1 + 2	.16	2.7

The combined score on the first two diagnostics taken under test like conditions does not seem to be predicted from the midterm percentage. As Diag 1 + 2 may be the very first mathematical type testing, other than the pretest, that students have encountered in several year, the testing situation maybe interferring with their problem solving abilities. Although the actual scores are not predicting the midterm percentage, the number of diagnostics taken does have an effect in predicting the midterm and final percentages. If the feedback from the diagnostics is working then the actual scores on these tests shouldn't be good predictors. Additional learning would be taking place as a result of the diagnostics, and this leads to improvement in the midterm and final performances. A great deal of useful learning may have resulted from the actual taking of the diagnostics. Our lack of control over the actual test taking situation does not allow us to make any further conclusions based on the number of diagnostics taken.

BEST COPY AVAILABLE

Differences in the change of administration of the diagnostics, formal testing (terms 1-3, 115 subjects) versus self-study (terms 4-8, 190 subjects) were tested in a one way analysis of variance. The null hypothesis of no differences could not be rejected ($p < .2902$). The testing situation does not seem to motivate students to a better performance than those who may have used the diagnostics as a self-study aid, but there are other variables confounded in this comparison and the implications of the lack of differences are less than clear.

Data for terms 1-3 also show that 3 people dropped after taking the first diagnostic, 7 people after the second dianostic, and 20 after they took the midterm. This is a drop out rate of 16%. Born and Whelan had a similar experience in a study of a Personalized System of Instruction with a beginning undergraduate psychology class (Psychological Record, 1973). There drop out rates were 3 to 5 times (14 to 25%) as great as more traditionally taught courses. They went on to show that generally the students with poorer academic records tended to drop out and usually this was because they fell behind in their own self-pacing. There were two questions posed for further research from their study:

1. Would increased numbers of study units and more frequent testing push the slow students?
2. Would the increased number of study units produce an inflexible study routine?

These quesions can not be answered here either, but perhaps with graduate studerts it may not be necessary to provide the extra testing situations. Those that find it helpful will do it on their own as they seemed to have done in terms 4 through 8.

Additional test taking practice is gained through choosing to take the parallel midterm (the highest of the two midterms was used in calculating the final grade). A decision rule that was cited to help students determine

BEST COPY AVAILABLE

whether to take the second midterm is: If you received a score of 24 or greater, roughly 80%, on midterm I, then taking the second midterm is unnecessary. Four comparison groups were formed based upon this rule: 1) 100 subjects who took only the first midterm, but obtained scores greater than or equal to 24, 2) 57 subjects who took both midterms and scored greater than or equal to 24 on the first midterm and scored greater than or equal to 24 on the first midterm, 3) 19 subjects who took the first midterm and scored below 24, and 4) 129 subjects who scored below 24, and took both midterms. A one way ANOVA allows the rejection of the null hypothesis of no differences between the four levels ($p < .0001$). Examination of the contrasts for all the dependent variables shows the same significance pattern. There are differences between groups 1 and 4 and between 2 and 4 as would be expected. The differences between groups 3 and 4 are, however, nonsignificant, but the pattern is in the right direction. And this difference is a whole grade difference. The difference between the scores of the second midterm and the first midterm taken is labeled Diff. The observed cell means for the four groups are:

Group	Midterm I	Midterm II	Diff	Final	Percent
1	89.33	0	0	82.19	86.12
2	86.49	85.55	-.94	80.05	85.10
3	69.17	0	0	61.72	69.37
4	61.68	78.28	16.61	65.12	73.50

Comparing the second midterm with the first, the people in group 4 show an average improvement of 17%. Group 4's final percentage would have probably been much lower if they had decided not to take the second midterm. Those in group 3 may have denied themselves an opportunity to put their midterm and final in the 80% category. The soundness of the decision rule seems to be supported by this analysis, and additional practice for those

BEST COPY AVAILABLE

with lower scores to help them achieve better seems in order.

The higher drop out rates within master model classes was mentioned in conjunction with the pretest. A more thorough discussion of the more general question: Does the extended time, granting of deferreds, tend to improve performance?, requires more attention here. Specifically, we wanted to know whether those who took longer to finish did better in performance than those who completed within one term? An ANOVA on 251 students who completed in one term versus 54 students who took 2 to 6 terms to complete found significant differences ($p < .0001$). These differences between both levels were significant for the three dependent variables: midterm, final, and percent. The observed cell means are:

Group	Midterm	Final	Percent
1	85.75	75.79	81.37
2	76.37	60.87	70.61

Those who defer do poorer by almost 10 percentage points on the midterm, which is taken prior to the granting of the deferred grade but end up worse by being 15 points lower on the final. We cannot tell whether they did gain in competencies with the extra time, ie closed the gap, or if they seemed only to be postponing an inevitable lower performance.

After the first term of the study in which 50% of those enrolled deferred, the deferral rate stabilized to an average of 10% per term. The drop out rate in the first year was 17%, and in the second year it had decreased to 14% (the second year students still have 1 year in which to complete their university requirements to make up the deferred). Although the requirements for attaining a deferred grade have not changed, the word may have been spread by other students that it really didn't work out to their benefit to take a deferred.

BEST COPY AVAILABLE

Significant differences were found in three separate ANOVAS for class (masters or doctoral), status (completed or deferred) and sex (males or female). A multivariate ANOVA with the following groups was done:

Group	Class	Status	Sex	Number
1	Masters	Completed	Male	38
2	Masters	Completed	Female	54
3	Masters	Deferred	Male	10
4	Masters	Deferred	Female	5
5	Doctoral	Completed	Male	77
6	Doctoral	Completed	Female	29
7	Doctoral	Deferred	Male	10
8	Doctoral	Deferred	Female	8

All interactions, those between class by status by sex, class by status, status by sex, and class by sex were not significant. The hypothesis for the main effects due to sex ($p < .0508$), to class ($p < .0027$), and to status ($p < .0001$) can be rejected. A similar pattern of differences exists for the dependent variables midterm, final, and percent for the two contrasts, the grand mean and status. The midterm is also significant for sex. The observed cell means are:

Group	Midterm (%)	Final (%)	Percent (%)
1	85.65	75.11	80.21
2	85.91	76.63	81.78
3	81.73	61.52	73.07
4	76.12	67.22	73.51
5	84.10	70.59	78.45
6	83.29	75.51	80.95
7	75.31	58.41	70.82
8	61.46	54.26	63.96

A surprising conclusion from the above table is that the masters students did better for every cell than did the students in a doctoral program. One viable explanation is that perhaps those who are in masters programs have more of a choice in selecting beginning statistics whereas the same course is usually a requirement for the doctoral student. The masters students maybe also thinking of entering a doctoral program, and working extra hard to impress the admissions committees with a good record

BEST COPY AVAILABLE

in statistics. The doctoral candidates maybe more deeply involved in other professional interests that tend to take up more of their study time. Remember that all of this is only speculation, nothing causal can be proposed here.

The males did significantly better than the females on the midterm (exception: group 1 and 2 where the means are almost identical) while the females did better on the final than the males (this difference was not significant, but in the right direction). These overall differences average out in the percent used for determining the grade, so they seem to be an unimportant aspect for further consideration.

The status differences, completed versus deferred, were already discussed in this paper. Again those who completed did significantly better on the midterm, final, and percent than those who deferred.

Finally, some over all comparisons involving the frequencies for classes in a year prior to the course (the more traditional class, year 1) under direct study, 869 for the two years of the study, years 2 and 3, and two terms after the study, still retaining many of the practices described in this study, year 4) can be examined for trends. Specifically, we wanted to know if there was an upward trend in the final percent attained across years. The goal of most mastery model courses is to produce 80% mastery for all or almost all students, this was not true here. However, it is interesting to see what did happen. Each percentage range in the following table corresponds to the grades assigned (1.5-4.0).

Year	50	50-64	65-69	70-84	85-89	90-100	Deferred
1	1	5	3	48	27	18	20
2	3	8	7	45	19	19	14
3	1	12	1	44	20	20	13
4	4	4	3	48	25	17	20

There is very little difference in each column of the above table when comparing over the years of the study, the prior class, and year 4. No matter what strategies were implemented, overall performance varied little. One implication, supported by our analyses using the pretest, is that improvement on entry skills would most likely have had the greatest chance of improving performance. We have no way of assessing whether any desireable affective changes occurred that could be attributed to the modified mastery strategies.

In conclusion, the most important results from this study are:

1. The pretest is a good predictor of performance in the class.
2. The number of diagnostics taken contributes to improved performance.
3. Scores on the second midterm almost always increase.
4. Increased time, ie the granting of deferrals, may not tend to improve performance.
5. Masters students always do better than doctoral students.
6. Overall performance, comparing classes prior to and after implementing the modified mastery model strategies, is about the same.

The results of this research will be used to improve advice to students taking the course in the future and to making changes within the current practices of the program in beginning statistics. Hopefully, the knowledge gained on these issues and strategies will be helpful and contribute to further work in the area of teaching beginning statistics.

BEST COPY AVAILABLE

BEST COPY AVAILABLE

APPENDIX

Colleges	Completed	Deferred	Total
Agriculture and Natural Resources	6	0	6
Business	16	0	16
Human Ecology	30	2	32
Natural Science	14	0	14
Education	208	43	251
Communication	8	2	10
Social Sciences	16	2	18
Missing Information	7	5	12
Total	305	54	359

Table 1

Class	Completed	Deferred	Total
Seniors	255	54	309
Graduate	50	0	50
Missing Information	0	0	0
Total	305	54	359

Table 2

Sex	Completed	Deferred	Total
Males	136	32	169
Females	98	13	111
Missing Information	71	10	81
Total	305	54	359

Table 3

COURSE DESCRIPTION

Topics for the first half of the course covers set theory, probability, probability models, and discrete and random probability distributions. The second half concentrates on the basic concepts of statistical inference, estimation and confidence intervals, hypothesis testing, large and small sample theory, simple correlation and linear regression techniques, three uses of the chi square test, and some properties of the binomial population.

REFERENCES

Airasian, P.W. An Application of a Mastery Learning Strategy, Psychology in the Schools, 1973, 130-134.

Block, J.H. Operating Procedures for Mastery Learning, In J.Block (Ed.) Mastery Learning: Theory and Practice. New York, Rhinehart and Winston, 1971, pp 64-76.

Bloom, B.S. Mastery Learning, In J.Block (Ed.) Mastery Learning: Theory and Practice. New York, Rhinehart and Winston, 1971, pp 47-63.

Born, D.G. and Whelan, P. Some Descriptive Characteristics of Student Performance in PSI and Lecture Courses, The Psychological Record, 1973, 23, 145-152.